On Evaluation of Recharge Model Uncertainty: a Priori and a Posteriori

Ming Ye*, Karl Pohlmann, Jenny Chapman, and David Shafer
Division of Hydrologic Sciences and Engineering, Desert Research Institute, Nevada System of Higher Education, 755 E. Flamingo Road, Las Vegas, NV 89119, United States
** Corresponding Author*

Abstract: *Hydrologic environments are open and complex, rendering them prone to multiple interpretations and mathematical descriptions. Hydrologic analyses typically rely on a single conceptual-mathematical model, which ignores conceptual model uncertainty and may result in bias in predictions and under-estimation of predictive uncertainty. This study is to assess conceptual model uncertainty residing in five recharge models developed to date by different researchers based on different theories for Nevada and Death Valley area, CA. A recently developed statistical method, Maximum Likelihood Bayesian Model Averaging (MLBMA), is utilized for this analysis. In a Bayesian framework, the recharge model uncertainty is assessed, a priori, using expert judgments collected through an expert elicitation in the form of prior probabilities of the models. The uncertainty is then evaluated, a posteriori, by updating the prior probabilities to estimate posterior model probability. The updating is conducted through maximum likelihood inverse modeling by calibrating the Death Valley Regional Flow System (DVRFS) model corresponding to each recharge model against observations of head and flow. Calibration results of DVRFS for the five recharge models are used to estimate three information criteria (AIC, BIC, and KIC) used to rank and discriminate these models. Posterior probabilities of the five recharge models, evaluated using KIC, are used as weights to average head predictions, which gives posterior mean and variance. The posterior quantities incorporate both parametric and conceptual model uncertainties.*

Ⅰ. INTRODUCTION

Hydrologic analyses are commonly based on a single conceptual-mathematical model. Yet hydrologic environments are open and complex, rendering them prone to multiple interpretations and mathematical descriptions. This is true regardless of the quantity and quality of available hydrologic information and data. Focusing on only one conceptual-mathematical model may lead to a Type I model error, which arises when one rejects (by omission) valid alternative models. It may also result in a Type II model error, which arises when one adopts (fails to reject) an invalid conceptual-mathematical model. Indeed, critiques of hydrologic analyses, and legal challenges to them, typically focus on the validity of the underlying conceptual (and by implication mathematical) model. If a proposed model is found to be severely deficient, hydrologic analysis based on the single model may damage professional credibility of the work; result in the loss of a legal contest; and lead to adverse environmental, economic and political impacts ([1-2]).

The need to properly assess conceptual model uncertainty has motivated the recent development of a Maximum Likelihood Bayesian Model Averaging method (MLBMA) [3-4]. MLBMA is being applied in our study to assess conceptual model uncertainty in the Death Valley Regional Flow System (DVRFS) model, developed by the U.S. Geological Survey [5] to simulate the regional flow system in southwest Nevada and southeast California. This area includes the U.S. Department of Energy proposed Yucca Mountain nuclear repository, the nation's first long-term permanent geologic repository of spent nuclear fuel and high-level radioactive waste.

Our study is focused on assessing conceptual model uncertainty due to five alternative recharge models listed in Table 1: (1) the Maxey-Eakin (ME) model [6], (2) two distributed parameter watershed (DPW) models, one with and one without a runon-runoff component [7], and (3) two chloride mass balance (CMB) models, each with different zero-recharge masks, one for alluvium and one for both alluvium and elevation [8]. These five models are based on different methodologies for estimating recharge and have different levels of complexity, and they all have been used for groundwater modeling in Nevada. Recharge estimates of the five models are plotted in Figure 1, and they are significantly different. A large amount of conceptual model uncertainty exits in the recharge models [9]. Since recharge significantly affects modeled groundwater flow paths and travel times, it is important to evaluate the recharge model uncertainty and quantify its propagation through the groundwater modeling process.

Using MLBMA, we assess the recharge model uncertainty *a priori* and *a posteriori*. The terms "*a priori*" and "*a posteriori*" refer primarily to how or on what basis our assessment is conducted. An assessment is conducted *a priori* if it is based on prior information without calibrating the regional flow model (of which the recharge model is a component) against site observations (e.g., hydraulic head and groundwater flux). The prior information includes assessment of model uncertainty from a similar site and/or expert judgments based on one's professional experience. An assessment is conducted *a posteriori* when the regional flow model (of which the recharge model is a component) is calibrated against site observations. In MLBMA, assessments *a priori* and *a posteriori* are quantified by prior and posterior model probabilities, as discussed below.

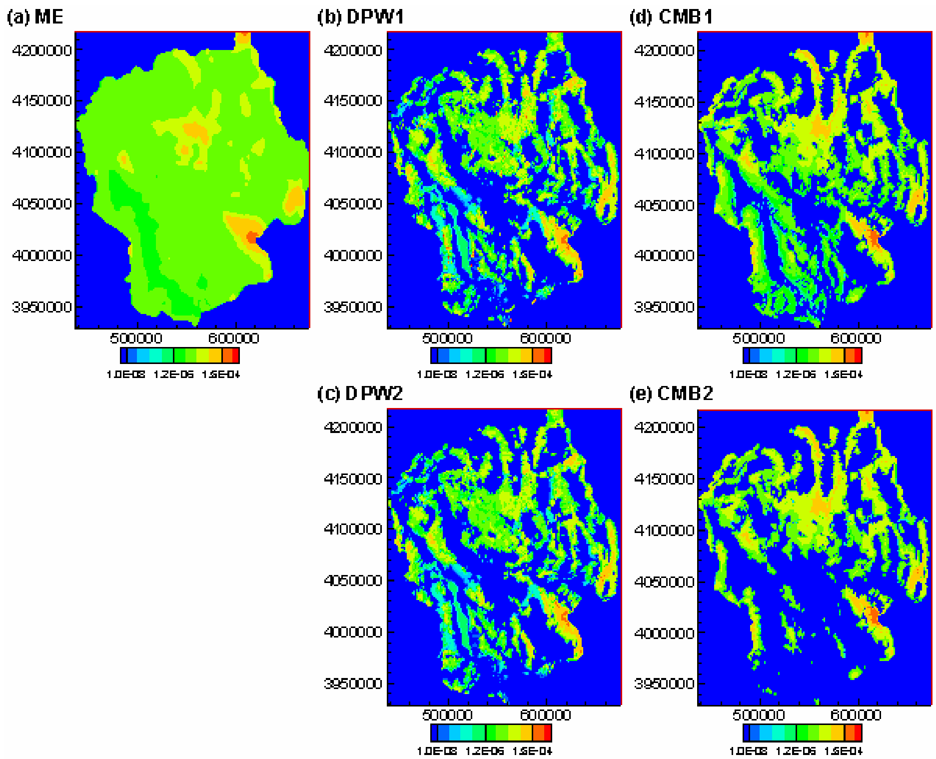

Figure 1. Illustration of the five recharge models: (a) ME, (b) DPW1, (c) DPW2, (d) CMB1, and (e) CMB2.

Table 1. Abbreviation and Description of the Five Recharge Models.

Models	Model Description
ME	Maxey-Eakin model
DPW1	Distributed parameter watershed model with runon-runoff component
DPW2	Distributed parameter watershed model without runon-runoff component
CMB1	Chloride mass balance model with fluvial mask
CMB2	Chloride mass balance model with fluvial and elevation masks

Ⅱ. MAXIMUM LIKELIHOOD BAYESIAN MODEL AVERAGING (MLBMA)

To render our paper complete and self-contained, we start with a brief description of MLBMA; for additional details the reader is referred to [3-4]. If Δ is the desired predicted quantity given a set of K alternative models, then its posterior distribution, given a discrete set $\mathbf{D}$ of site data, is

$$p\left(\Delta|\mathbf{D}\right)=\sum_{k=1}^{K}p\left(\Delta|M_k,\mathbf{D}\right)p\left(M_k|\mathbf{D}\right) \tag{1}$$

where $p(\Delta|M_k,\mathbf{D})$ is the posterior distribution of Δ under model M_k and $p(M_k|\mathbf{D})$ is posterior probability of M_k. With consideration of parametric and conceptual model uncertainty, mean and variance of Δ are

$$E\left[\Delta|\mathbf{D}\right]=\sum_{k=1}^{K}E\left[\Delta|\mathbf{D},M_k\right]p\left(M_k|\mathbf{D}\right) \tag{2}$$

$$\begin{aligned}Var\left[\Delta|\mathbf{D}\right]=&\sum_{k=1}^{K}Var\left[\Delta|\mathbf{D},M_k\right]p\left(M_k|\mathbf{D}\right)\\&+\sum_{k=1}^{K}\left(E\left[\Delta|\mathbf{D},M_k\right]-E\left[\Delta|\mathbf{D}\right]\right)^2p\left(M_k|\mathbf{D}\right)\end{aligned} \tag{3}$$

where $E\left[\Delta|\mathbf{D},M_k\right]$ and $Var\left[\Delta|\mathbf{D},M_k\right]$ are mean and variance of Δ under model M_k due to uncertainty of parameters associated with M_k. The weight $p(M_k|\mathbf{D})$ used to average model predictions and corresponding predictive variance is posterior model probability of model M_k, evaluated using Bayes' rule

$$p\left(M_k|\mathbf{D}\right)=\frac{p\left(\mathbf{D}|M_k\right)p\left(M_k\right)}{\sum_{l=1}^{K}p\left(\mathbf{D}|M_l\right)p\left(M_l\right)} \tag{4}$$

where $p(\mathbf{D}|M_k)$ is likelihood of model M_k (a measure of consistency between model predictions and site observations $\mathbf{D}$) and $p(M_k)$ is prior probability of M_k. Estimating prior model probability will be discussed in detail in Section Ⅲ.The posterior model probability is conditioned on site observations explicitly and prior information implicitly. According to [4], equation (4) can be approximated as

$$p\left(M_k|\mathbf{D}\right)\approx\frac{\exp\left(-\frac{1}{2}KIC_k\right)p\left(M_k\right)}{\sum_{l=1}^{K}\exp\left(-\frac{1}{2}KIC_l\right)p\left(M_l\right)} \tag{5}$$

where KIC is Kashyap information criterion defined as [10]

$$KIC_k=(N-N_k)\ln\sigma_k^2-N_k\ln(e\cdot 2\pi)+\ln|\mathbf{X}_k^T\boldsymbol{\omega}\mathbf{X}_k| \tag{6}$$

where N is number of calibration data $\mathbf{D}$, N_k is number of parameters $\boldsymbol{\theta}_k$ associated with model M_k, e is the natural number, $\boldsymbol{\omega}$ (the same for all models) is weight matrix associated with calibration data $\mathbf{D}$, and $\mathbf{X}$ is sensitivity matrix with element $X_{k,ij}=\partial D'_{k,i}\big/\partial\theta_{k,j}$ evaluated at maximum likelihood parameter estimates, $\hat{\boldsymbol{\theta}}_{k,j}$ ($D'_{k,i}$ being predictions at locations of D_i by model M_k). $\hat{\boldsymbol{\theta}}_{k,j}$ can be estimated using maximum likelihood (or, equivalently, generalized least square) methods, which also gives the calculated error variance, σ_k^2,

$$\sigma_k^2=\frac{\mathbf{e}_k^T\boldsymbol{\omega}\mathbf{e}_k}{N-N_k}=\frac{WSSR_k}{N-N_k} \tag{7}$$

where $\mathbf{e}=\mathbf{D}-\mathbf{D}'$ is residual and $WSSR_k$ is weighted sum of squared residual of model M_k. All the quantities above can be estimated based on results of model calibration using common software such as MODFLOW2000 [11].

Ⅲ. EVALUATE RECHARGE MODEL UNCERTAINTY: A PRIORI

Conceptual uncertainty of the five recharge models is first evaluated, *a priori*, using prior probabilities of the models. Prior model probability is interpreted by [4] as subjective values reflecting the analyst's (or a group of analysts') belief about the relative plausibility of each model (or a group of models) based on its apparent (qualitative, *a priori*) consistency with available knowledge and data. The analyst's perception, degree of reasonable belief [12], or confidence [13] in a model is ideally based on expert judgment, which is considered by [14] as the basis of conceptual model development. Hence we view integrating expert judgment in MLBMA (by specifying subjective prior probabilities) to be a strength rather than a weakness. According to this view, the models included in the model set must be those (and only those) that experts consider being of potential relevance to the problem at hand. Given a set of alternative models, their prior probabilities sum up to one,

$$\sum_{k=1}^{K}p\left(M_k\right)=1 \tag{8}$$

This implies that all possible models of relevance are included in the model set (collective exhaustiveness), and that all models in the set differ from each other sufficiently to be considered mutually exclusive (the joint probability of two or more models being zero).

Following the process suggested by [15], an expert elicitation is conducted to elicit professional judgments from seven experts on uncertainty of the five recharge models. Elicited prior probabilities of the five models are plotted in Figure 2, which shows that the maximum and minimum prior probabilities are 45% and 5%, respectively. Models ME, DPW1, and CMB2 are the three most plausible models, and do not receive the minimum prior probability from any expert. Although the experts evaluate the models from various aspects (e.g., model assumptions and sensitivity of model predictions to mode parameters), no experts place more than 50% prior probability on any model.

The prior model probabilities are aggregated using simple averaging i.e.,

$$P_k=\frac{1}{NE}\sum_{i=1}^{NE}P_{ik} \tag{9}$$

where NE=7 is number of experts and P_{ik} is the prior probability expert i assigns to alternative model M_k,. The aggregated prior model probabilities are plotted in Figure 3. Models DPW1 and DPW2 have the largest and smallest probability, respectively, since the experts regard that including the runon-runoff component is more realistic. Probability of the model CMB2 is larger than that of CMB1, since experts regard that elevation mask is an important feature in recharge estimation. Model ME is ranked as the second most plausible model and has prior model probability of 25%, although this model is the simplest one and its recharge estimation is significantly different from that of other four models. Although prior probabilities given by each expert are significantly different (Figure 2), the aggregated probabilities are more or less uniform, considering that the equally likely prior probability is 20%. The largest deviation from the equally likely prior probability is only 10% for model DPW1. This manifests the inherent uncertainty in the recharge models, since they are developed independently based on solid physical principles and assumptions, calibrated with site measurements, and have all been applied to water resource management in Nevada. Since none of the models dominates over other models and all models have prior model probabilities larger than 5%, there is no justification to select one model and discard others, *a priori*.

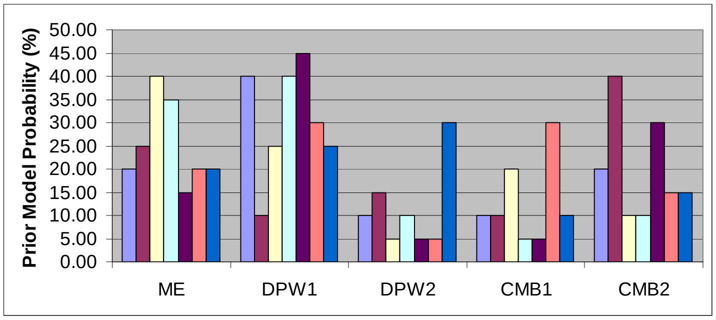

Figure 2. Column chart of prior probabilities of the five models given by seven experts. Columns of each model represent elicited prior model probability from one expert. Model names are explained in Table 1.

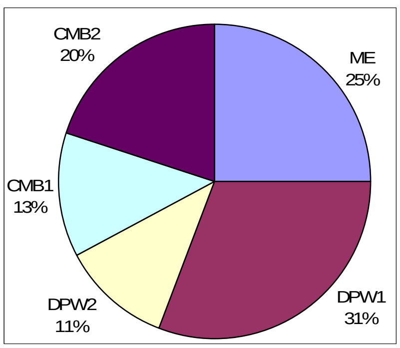

Figure 3. Prior probabilities of the five recharge models obtained through an expert elicitation. Model names are explained in Table 1.

Ⅳ. EVALUATE RECHARGE MODEL UNCERTAINTY: A POSTERIORI

Recharge model uncertainty is assessed, *a posteriori*, by maximum likelihood model calibration against site observations. Results of model calibration are used to estimate model likelihood $p(\mathbf{D}|M_k)$, which, in turn, is used to evaluate posterior model probability $p(M_k|\mathbf{D})$ in (4). Whereas prior model probabilities must in our view remain subjective, the posterior model probabilities are modifications of these subjective values based on an objective evaluation of each model's consistency with available data.

Ⅳ.A. Model Calibration Using MODFLOW2000

Plausibility and uncertainty of each of the five recharge models is evaluated by calibrating the Death Valley Regional Flow System (DVRFS) model, of which the recharge model is a component. DVRFS was modeled by [5] using MODFLOW2000, and a three-dimensional hydrogeologic framework based on characterization of regional geology, hydrology, and hydrogeology. The recharge model used in DVRFS is DPW1 developed by [7]. Our study is to assess recharge model uncertainty in the modeling framework of DVRFS, without modifying its other components. DVRFS was calibrated using MODFLOW2000 against a total of 4,963 observations of head (2,227), head change (2,672), discharge (49), and constant-head flow (15). These observations are also used in our calibration.

Our calibration process, however, is different from that of DVRFS, which calibrated 55 model parameters, 23 in the steady-state model and 32 in the transient model. Our model calibration is based on the transient model only, since there is insufficient information to identify how the 23 parameters are calibrated in the steady-state model. In addition, only some of the 55 parameters are calibrated in our study, due to different purposes of our study. Specifically, 32 of the 55 parameters are calibrated for DPW1 and DPW2. The two models estimate precipitation (not recharge), which is converted to recharge within DVRFS by dividing the top model layer into five recharge zones. Recharge coefficients in two zones are calibrated against site observations. Since the other three models estimate recharge directly, recharge coefficients are not used and therefore only 30 parameters are calibrated. All other calibration parameters are the same as those used in DVRFS. Although MLBMA allows different models having different numbers of calibrated parameters, we intend to calibrate the same model parameters for all the recharge models so that model ranking and uncertainty analysis are on the same basis. In the same line, model calibration is conducted in the same manner for all the recharge models. Specifically, all the model calibrations use identical initial parameter values, convergence criterion, and other calibration variables such as parameter log transform and damping factors.

Model calibration results corresponding to the five recharge models are summarized in Table 2 and Figure 4. Table 2 lists WSSR (weighted sum of squared residuals) of the four kinds of observations, respectively, and total WSSR. WSSR of DVRFS is also listed for comparison. The table shows that, except for recharge model ME, the values of WSSR of the models are close and are lower than that of DVRFS. This is not surprising since our calibration is based on model calibration of DVRFS to a certain extent, and can be regarded as further calibration of DVRFS. The largest relative differences of WSSR occur for the observations of discharge and constant-head flow. Figure 4 plots some of the calibrated parameters whose values are noticeably different between the recharge models (values of other calibrated parameter are close). The values of calibrated parameters in DVRFS are also plotted for comparison. Although WSSR corresponding to the recharge models are similar, some parameter values are different, indicating different responses of the regional flow system simulation to the recharge models, which provides a basis for model discrimination. The largest difference of parameter values occurs for hydraulic conductivity of volcanic rock units (K3) such as K3BRU123 and K3CTM. While most of the parameters are within the parameter ranges given in [5], several parameter values exceed the ranges. This, however, is not surprising, since the ranges are based on limited information of site measurements. All values of the calibrated parameters are considered reasonable.

TABLE 2. Weighted Sum of Squared Residuals (WSSR) for all Kinds of Observations Corresponding to the Five Recharge Models and DVRFS.

Type of observation	Observaion Number	DVRFS	ME	DPW1	DPW2	CMB1	CMB2
Hydraulic head	2227	23083.22	26321.55	20030.92	20296.37	20215.87	19803.57
Head changes	2672	13348.08	11805.63	12599.57	12752.66	12372.07	12057.11
Discharge	49	637.64	2078.36	674.43	611.12	1001.34	1062.06
Constant-head flow	15	438.15	1520.28	296.94	350.56	863.24	641.49
Total	4963	37507.10	41725.82	33601.86	34010.71	34452.52	33564.23

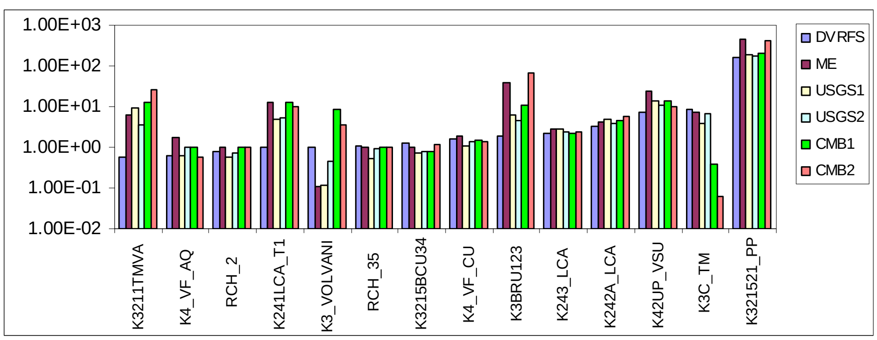

Figure 4. Comparison of values of some calibrated parameters corresponding to the five recharge models.

Ⅳ.B. Posterior Model Probability

Posterior model probabilities of the recharge models are calculated using equation (6) based on the model calibration results listed in Table 3, which also lists three information criteria (AIC, BIC, and KIC) commonly used to rank alternative models. AIC and BIC are evaluated via

$$AIC_k = -2\ln p\left(\mathbf{D}\middle|M_k\right) + 2N_k = N\ln\sigma_k^2 + N_k \qquad (10)$$

$$\begin{aligned} BIC_k &= -2\ln p\left(\mathbf{D}\middle|M_k\right) + N_k \ln N \\ &= N\ln\sigma_k^2 + N_k(\ln N - 1) \end{aligned} \qquad (11)$$

These information criteria rank alternative models not only based on their goodness-of-fit (as measured by WSSR) but also on the principle of parsimony, which states that a simple model (with lower number of parameters) is considered more plausible than a complex model if their predictions fit observations equally well. The three information criteria rank the five recharge models at almost the same order, with models ME, DPW2, and CMB1 ranked as the least plausible. In addition, the information criteria show that DPW1 and CMB2 are more plausible than DPW2 and CMB1, respectively, which is consistent with the results of the expert elicitation (Figure 3). Nevertheless, model ME is ranked as least plausible after model calibration. AIC and BIC rank CMB2 as the best model, while KIC ranks DPW1 as the best one. Inconsistency of model ranking given by different information criteria is not uncommon. Among the three criteria, KIC is favored since it incorporates quality of data used for model calibration [3] and can yield more reliable model rankings in various circumstances [e.g., 4].

TABLE 3. Quality Criteria, Ranking, And Prior/Posterior Probabilities Associated With The Five Recharge Models.

	ME	DPW1	DPW2	CMB1	CMB2
N_k	30	32	32	30	30
WSSR	41726	33602	34011	34453	33564
ln|F|	360	346	344	349	346
AIC	10627	9556	9616	9676	9547
Rank	5	2	3	4	1
BIC	10822	9765	9825	9872	9742
Rank	5	2	3	4	1
KIC	10808	9718	9775	9852	9720
Rank	5	1	3	4	2
$p(M_k)$	25%	30%	11%	13%	20%
$p(M_k\|\mathbf{D})$	0	83.25%	0	0	16.75%
$p(M_k)$	20%	20%	20%	20%	20%
$p(M_k\|\mathbf{D})$	0	76.82%	0	0	23.18%

Posterior model probabilities are evaluated using (5) for two sets of prior model probabilities. One set has informative priors obtained from the expert elicitation and the other one treats the five models equally likely. Regardless of prior probabilities, posterior probabilities of models ME, DPW2, and CMB2 are zero, indicating that they are implausible given the calibration data. This is so even though model ME received a relatively large prior probability from the experts. The effects of prior on posterior model probability is observed for models DPW1 and CMB2. Posterior probability of DPW1 decreases 6.43% when its prior probability decreases 10%. This results in a concomitant increase of 6.43% in the posterior probability of model CMB2, even though its prior probability does not change. Although it is expected that sensitivity of posterior to prior model probability diminishes as the amount conditioning (calibration) data increases, this study shows that, even with 4,963 observations, sensitivity to prior probability does not disappear. In this case, using informative prior model probability (obtained from expert elicitation in this study) may increase accuracy of model uncertainty assessment, as suggested in [16]. Note that just like prior probabilities, posterior probabilities are valid only in a comparative, not in an absolute, sense. They are conditional on the choice of models, calibration data, and prior information used to estimate prior model probabilities.

Ⅴ. BAYESIAN MODEL AVERAGING

Based on equations (2) and (3), Bayesian model averaging is used to yield the posterior mean and variance to incorporate both parametric and conceptual model uncertainty. The posterior mean represents the optimum prediction and the posterior variance measures the associated predictive uncertainty. Monte Carlo simulation is used to assess parametric uncertainty and estimate $E\left[\Delta\middle|\mathbf{D},M_k\right]$ and $Var\left[\Delta\middle|\mathbf{D},M_k\right]$ for model M_k. Multivariate normal distributions are used to generate 200 parameter realizations of the calibrated parameters. The mean of the normal distribution is the maximum likelihood parameter estimate $\hat{\boldsymbol{\theta}}_k$ of model M_k and the covariance matrix is $\sigma^2(\mathbf{X}_k^T\boldsymbol{\omega}\mathbf{X}_k)^{-1}$ [3, 11]. Estimation of posterior mean and variance using equations (2) and (3) is straightforward. Figure 5 plots mean head predictions corresponding to the five recharge models (Figures 5b-5f) and the MLBMA (Figure 5a) posterior mean head in the first (top) model layer at stress period 87 (1998, the last year of the transient model). The MLBMA mean head (Figure 5a) is an average of the mean heads for DPW1 and CMB2, since the other three models have zero posterior probabilities. Mean head contours of DPW1 and DPW2 are similar to each other, as are the contours of CMB1 and CMB2, owing to the similarity of the two pairs of recharge models. Since these four recharge models are different from ME, the contour of ME is different from the four contours in Figures 5c-5f. Figure 6 plots the cumulative distribution function (CDF) of the mean head and standard deviation of head predicted by models DPW1 and CMB2 and MLBMA for the entire simulation domain at stress period 87. For mean head predictions, the CDFs of DPW1, CMB2, and MLBMA are almost identical, due to the similarity of mean head predictions of DPW1 and CMB2 (shown in Figures 5b and 5f). Nevertheless, the standard deviation of head prediction of MLBMA is larger than that of models DPW1 and CMB2, since MLBMA considers both parametric and conceptual model uncertainty, while the two single models address only parametric uncertainty.

Ⅵ. CONCLUSIONS

This study assesses conceptual model uncertainty of five recharge models within the modeling framework of DVRFS, of which each recharge model is a component. Conceptual model uncertainty is first assessed, *a priori*, using expert judgment gathered from an expert elicitation. The experts placed higher probabilities on DPW1and CMB2 than DPW2 and CMB1, respectively. However, since the recharge models are developed by different researchers based on different theories, prior model probabilities elicited from the experts are around the average value of 20%. This indicates that one cannot select one model for predictions and discard all others *a priori*. Since prior information cannot fully assess conceptual model uncertainty, model calibration is needed to assess conceptual model uncertainty *a posteriori* based on observations of head and flow. DVRFS is used as the framework for numerical modeling and only its recharge component varies for different recharge models (other components remain the same). Based on model calibration results using MODFLOW2000, three information criteria (AIC, BIC, and KIC) are evaluated to rank the models. Model ranking of AIC and BIC are the same, but different from that of KIC. Consistent with results of expert elicitation, DPW1 and CMB2 are ranked more plausible than DPW2 and CMB1, respectively. However, as opposed to the results of expert elicitation, model ME is ranked as least plausible. This suggests the importance of uncertainty assessment *a posteriori*. Posterior model probabilities are evaluated using KIC, which is considered superior to AIC and BIC. Models ME, DPW2, and CMB1 have zero posterior probabilities. Sensitivity of posterior to prior probabilities for models DPW1 and CMb2 does not disappear, although 4,963 observations are used for model calibration. Note that posterior probabilities are valid only in a comparative, not in an absolute, sense. They are conditional on the choice of models, calibration data, and prior information used to estimate prior probabilities. Bayesian model averaging is conducted to estimate posterior mean and variance of head and flux. Posterior variance of MLBMA is larger than the variance of any single model, since conceptual model uncertainty is also addressed. Our research results can be extended to incorporate conceptual model uncertainty in flow path delineation, which can in turn be used to design networks for detection and monitoring of potential radionuclide transport in the saturated zone of the Death Valley Regional Flow System, where Yucca Mountain is located.

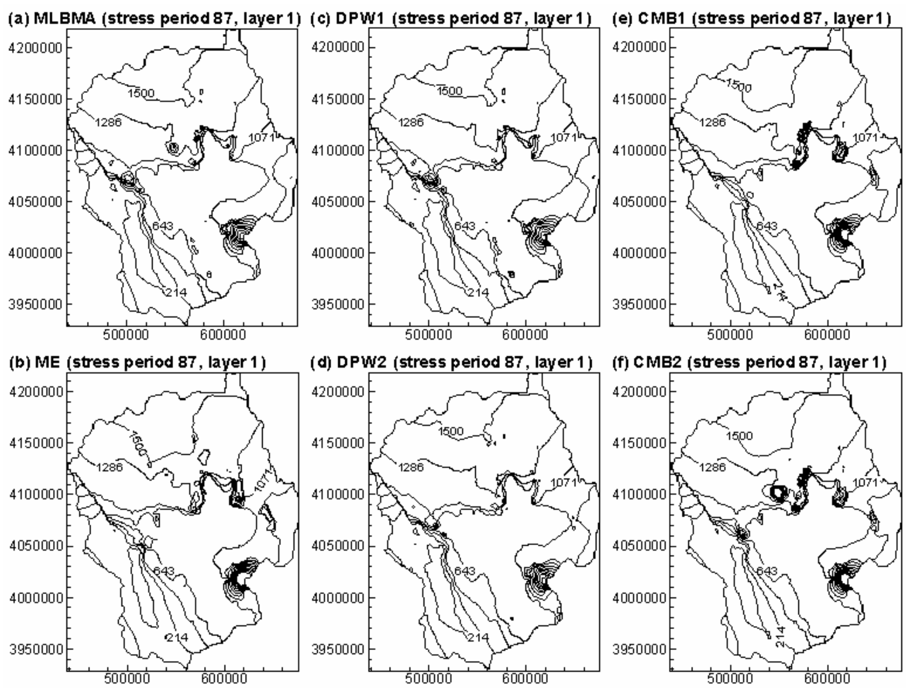

Figure 5. Mean head predicted by (a) MLBMA, (b) ME, (c) DPW1, (d) DPW2, (e) CMB1, and (f) CMB2 in the first (top) layer at stress period 87 (1998).

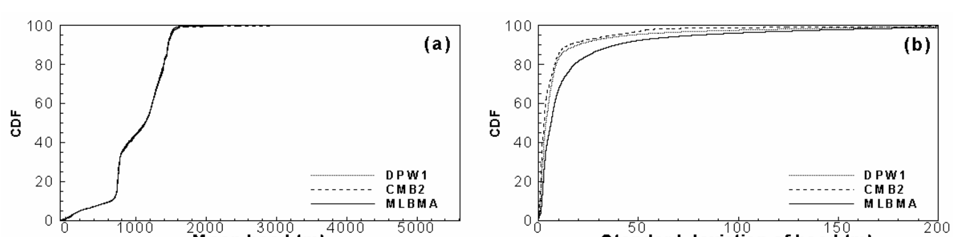

Figure 6. Cumulative distribution function (CDF) of mean head and standard deviation of head prediction over the whole simulation domain at stress period 87 (1998).

ACKNOWLEDGMENTS

Research is funded through contract DE-AC52-00NV13609 with the U.S. Department of Energy, National Nuclear Security Administration Nevada Site Office (NNSA/NSO). We appreciate Dr. Glendon Gee, Dr. Greg Pohll, Mr. Chuck Russell, Dr. Joe Hevesi, and Mr. Randy Laczniak for their participation in the expert elicitation.

REFERENCES

[1] NATIONAL RESEARCH COUNCIL, "Conceptual Models of Flow and Transport in the Fractured Vadose Zone", National Academy Press, Washington, DC (2001).

[2] S.P. NEUMAN and P.J. WIERENGA, "A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites", NUREG/CR-6805, U.S. Nuclear Regulatory Commission, Washington, D.C. (2003).

[3] S.P. NEUMAN, "Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models, Stochastic Environmental Research and Risk Assessment", 17(5), 291-305, DOI: 10.1007/s00477-003-0151-7 (2003).

[4] M. YE, S.P. NEUMAN, and P.D. MEYER, "Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff", Water Resour. Res., 40, W05113, doi:10.1029/2003WR002557 (2004).

[5] W.R. BELCHER (ed). "Death Valley regional ground-water flow system, Nevada and California – Hydrogeologic framework and transient ground-water flow model", U.S. Geological Survey Scientific Investigation Report 2004-5205 (2004).

[6] G.B. MAXEY and T.E. EAKIN, "Ground water in White River Valley, White Pine, Nye, and Lincoln Counties, Nevada". No. 8, State of Nevada Office of the State Engineer prepared in cooperation with the United State Department of the Interior Geological Survey, Carson City, Nevada (1949).

[7] J.A. HEVESI, A.L. FLINT, and L.E. FLINT, "Simulation of Net Infiltration and Potential Recharge Using a Distributed Parameter Watershed Model for the Death Valley Region, Nevada and California". Water-Resources, Investigations Report 03-4090. Sacramento, CA: U.S. Geological Survey (2003).

[8] C.E. RUSSELL, and T. MINOR, "Reconnaissance Estimates of Recharge Based on an Elevation-dependent Chloride Mass-balance Approach", DOE/NV/11508-37. Prepared for the U.S. Department of Energy, National Nuclear Security Administration Nevada Operations Office. Las Vegas, NV: Desert Research Institute (2002).

[9] K. REHFELDT (ed), "Hydrologic Data For The Groundwater Flow And Contaminant Transport Model Of Corrective Action Units 101 And 102: Central And Western Pahute Mesa, Nye County, Nevada", Stoller-Navarro Joint Venture (2004).

[10] R.L KASHYAP., "Optimal choice of AR and MA parts in autoregressive moving average models", IEEE Trans. Pattern Anal. Mach. Intel. PAMI, 4(2), 99-104 (1982).

[11] M.C. HILL, E.R. BANTA, A.W. HARBAUGH, and E.R. ANDERMAN, MODFLOW-2000, "The U.S. Geological Survey modular groundwater model – User guide to the observation, sensitivity, and parameter estimation processes and three post-processing programs", USGS Open-File Report 00-184, Reston, Virginia: USGS (2000).

[12] H. JEFFREYS, "*Scientific Inference*" (2nd ed.), Cambridge, U.K., Cambridge University Press, (1957).

[13] E. ZIO, and G.E. APOSTOLAKIS, "Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories", *Reliability Engineering and System Safety*, 54, 225-241 (1996).

[14] J. BREDEHOEFT," The conceptualization model problem-surprise", Hydrogeol. J., 13, 37-46 (2005).

[15] R.L. KEENEY AND D. VON WINTERFELDT, "Eliciting probabilities from experts in complex technical problems", IEEE Transactions On Engineering Management, 38 (3): 191-201 (1991).

[16] M. YE, S. P. NEUMAN, P. D. MEYER, and K. POHLMANN (2005), "Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff", Water Resour. Res., 41, W12429, doi:10.1029/2005WR004260.